



Bridging operations research and machine learning for service cost prediction in logistics and service industries

Marco Boresta¹ · Diego Maria Pinto¹ · Giuseppe Stecca¹

Received: 12 May 2023 / Accepted: 21 March 2024 / Published online: 21 May 2024
© The Author(s) 2024

Abstract

Optimizing shared resources across multiple clients is a complex challenge in the production, logistics, and service sectors. This study addresses the underexplored area of forecasting service costs for non-cooperative clients, which is essential for sustainable business management. We propose a framework that merges Operations Research (OR) and Machine Learning (ML) to fill this gap. It begins by applying the OR model to historical instances, optimizing resource allocation, and determining equitable service cost allocations for each client. These allocations serve as training targets for ML models, which are trained using a combination of original and augmented client data, aiming to reliably project service costs and support competitive, sustainable pricing strategies. The framework's efficacy is demonstrated in a reverse logistics case study, benchmarked against two traditional cost estimation methods for new clients. Comparative analysis shows that our framework outperforms these methods in terms of predictive accuracy, highlighting its superior effectiveness. The integration of OR and ML offers a significant decision-support mechanism, improving sustainable business strategies across sectors. Our framework provides a scalable solution for cost forecasting and resource optimization, marking progress toward a circular, sustainable economy by accurately estimating costs and promoting efficient operations.

Keywords Machine learning · Cost allocation · Vehicle routing problem · Explainable artificial intelligence · Waste management · Fairness

1 Introduction and literature review

The use of information systems, mathematical models, and AI-based solutions has gained popularity in recent years for meeting the ambitious targets of the sustainability paradigm. This trend is particularly noticeable in the manufacturing and service sectors, where there

✉ Diego Maria Pinto
diegomaria.pinto@iasi.cnr.it

Marco Boresta
marco.boresta@iasi.cnr.it

Giuseppe Stecca
giuseppe.stecca@iasi.cnr.it

¹ Istituto di Analisi dei Sistemi ed Informatica del CNR, Via dei Taurini 19, 00185 Roma, Italy

is a clear focus on achieving economic growth and development while limiting the adverse effects on the environment. Organizations employ innovative management approaches to attain these objectives, recognizing the intrinsic connection between sustainable efficiency and profitability. Approaches vary by industry, and when processes involve multiple customers, the optimal strategy is to share production or service capacity among them, thus limiting the use of resources. This approach requires solving a variety of decision-making problems, such as (i) optimizing resource allocation across multiple customers, (ii) allocating a fair portion of the overall production/service cost to each of them, (iii) forecasting the expected cost of serving a new customer to ensure competitiveness and profitability.

1.1 Research questions and contributions

Considering the set of decision problems (i), (ii), and (iii), the research question of this paper is how to predict a fair and optimal cost for serving each client within a group of non-cooperative clients served with limited resources. This scenario presents significant challenges, including issues related to data uncertainty and fairness, since the cost allocation problem is fundamentally concerned with fairness.

Accordingly, we aim to expand the research on this set of problems by leveraging OR and ML techniques. The objective is to enhance decision-making in the domain of sustainable business management, with particular emphasis on the logistics and service sectors.

To tackle the stated research question, this paper introduces a novel integration of OR and ML within a decision-support framework. This integration effectively merges the operational efficiency of OR with the predictive power of ML in service cost estimation, considering the different characteristics of clients. The result is a versatile decision-support tool that facilitates rapid and precise cost estimation, thereby enhancing margin assessments and enabling more informed, client-specific cost allocations. Consequently, it offers a practical solution to the challenges of low margins and uncertainties commonly encountered in Circular Economy (CE) processes across various service sectors.

In our framework, we begin with generating instances of the resource allocation problem, which are then modeled and solved using OR techniques, after which costs are allocated to each client. The next phase involves training a Machine Learning (ML) model to accurately predict these client-specific costs. A dedicated section in the paper provides a detailed overview of the framework, elucidating each of these distinct phases.

1.2 Case study

This research's real-world application focuses on a logistics scenario where capacity-sharing solutions ensure sustainability, even in the absence of collaboration among participants. These participants are the clients sharing the transport capacity of a logistic company that provides pickup and delivery solutions, and their varying locations, visiting time windows, and service variability may serve as valuable features for the client cost forecasting model. In the proposed application, the ML model aims to allow vendors to forecast the cost of serving clients, enabling fair and competitive pricing while ensuring business sustainability and profitability. Moreover, the model is trained on historical data and the optimal allocation of resources can enable vendors to offer reliable contracts based on expected service costs. Sales managers could use cost forecasts when negotiating a new contract with a prospect or (re-)negotiating contracts with clients, charging them their expected portion of the overall

service cost. In particular, contract re-negotiation is a consolidated practice in the CE and recycling sector, where secondary raw material prices vary.

A dedicated section provides a detailed presentation of this use-case application. The particular Reverse Logistics (RL) problem is modeled as a variant of the vehicle routing problem with pickup and deliveries. In particular, a truck can transport 1 or 2 waste containers (full or empty) between clients, depots, and recycling plants. Among the specific features of this model, there are heterogeneous trucks in terms of operating and emissions costs, specific hauling capacities, and non-linear costs modeling truck driver costs. In fact, there is a two-part tariff comprised of a fixed fee plus a cost per driving hour. Moreover, an additional overnight cost can be incurred, potentially allowing trucks to take multiple driving shifts to reach remote locations. Appendix A provides a detailed explanation of the RL formulation.

1.3 Policy implication for circular economy

Circular economy business models are driven by different policies. Generally speaking, policies aim at closing the loop for material flows while increasing material efficiency. Milios (2018) identifies three main policy areas: (1) policies for reuse, repair, and remanufacturing; (2) green public procurement and innovation procurement; and (3) policies for fostering the Secondary Raw Materials (SRM) market. In this research work, we propose a decision-support framework producing lower-cost solutions for closing the CE loop on materials flows while increasing the profitability of the collection process, thereby enhancing its business feasibility. Concerning the policy areas defined by Milios (2018), the proposed framework affects areas (1) and (3). Indeed, this innovative approach is instrumental for businesses operating in dynamic and complex environments such as those found in the CE, offering a practical solution to the challenges of low margins and uncertainties commonly encountered in this sector. This is demonstrated by the application of the framework to the considered use case, specifically addressing the collection of SRM. A dedicated section presents the potential of the framework through numerical experiments using real-world instances given by the use-case provider. CE must be economically sustainable to be a realizable target for business companies. In this regard, the developed framework can obtain a profitable collection of material thanks to the integration of OR and ML, as demonstrated in the analyzed use case.

Furthermore, this framework applies to any OR model that allocates capacity or resources among multiple non-cooperative clients. It is particularly beneficial for logistics solutions, where capacity sharing serves as a viable pathway towards sustainability and the realization of CE principles.

1.4 Literature review

This section presents a literature review that spans the key thematic areas pertinent to our study: resource allocation and related cost sharing, Circular Economy and related SRM and WM processes, and the integration of Operations Research (OR) and Artificial Intelligence (AI) methodologies in Decision Support Systems devoted to these contexts.

In Yu et al. (2015), a set of firms have the choice of either operating their own production/service facilities or investing in a facility that is shared; the problem is formulated as a cooperative game, with cost allocation under either the first-come, first-served policy or an optimal priority policy. In Liu et al. (2022), firms decide on the allocation of demand from different sources to different facilities to minimize delay and service-fulfillment costs. In the realm of logistics, Verdonck et al. (2016) introduce a novel approach for exploiting horizon-

tal carrier collaboration. This involves the sharing of distribution centers among partnering organizations. This approach enables transport companies to cooperate, increasing their efficiency levels, for example, by exchanging orders or vehicle capacity. To ensure cooperation sustainability, the same source addresses how collaborative costs need to be allocated fairly to the different participants, and they analyze the effects of different cost allocation techniques with numerical experiments. Thus, also in the aforementioned work, resources are shared across firms that decide to operate collectively as a coalition. Moreover, they highlight the importance of a fair cost allocation of shared resources. Angelelli et al. (2022) explore resource sharing in horizontal collaboration among taxis within Demand-Responsive Transport (DRT) settings. Their work involves defining and solving optimization problems that aim to minimize the total service cost while ensuring a balanced workload among operators.

The survey conducted by Barros et al. (2021) maps the impact of Circular Economy practices across various business areas. In particular, several articles demonstrate how circular economy contributes to improving resource efficiency and sustainable economic development and there is a positive correlation between objectives seeking to lower the environmental burdens and the improvement of the economic aspects of their operations revenue profit. The authors identify nine business areas: strategic planning, cost management, supply chain management, quality management, environmental management, process management, logistics and reverse logistics, service management, and research and development, and show how CE impacts business performance. Another extensive survey by Lai and Wong (2012) discusses the positive impact of green supply chain practices on both the operational and environmental performances of the studied Chinese companies.

Fatimah et al. (2020) analyzes the relationship between CE, sustainable development goals, and digitization (Industry 4.0) in a case study. It is interesting to note how the maturity model applied in the paper put into evidence the importance of Decision Support Systems (DSS) to sustain the solution of technology options problems.

The work by Bag et al. (2021) shows an interesting example of how big data analytics and artificial intelligence can be used to draw relationships between circular economy practices, sustainable business, and institutional regulations. In the study, applied to a South African case, the methodology is used to estimate the value of relations among institutional pressures and tangible resources employed for circular economy practices.

In the specific context of Waste Management (WM) within logistics, the works of Stecca and Kaihara (2021) and Mancini and Gansterer (2022) offer innovative strategies for operational optimization. Stecca and Kaihara (2021) introduces a negotiation-based scheme aimed at leveraging client aggregation to boost service efficiency. Complementarily, Mancini and Gansterer (2022) explores the strategic use of clustering and corridor creation as methods for client aggregation, showcasing their potential to significantly enhance operational efficiency in WM logistics.

Considering the CE and its closed loop of SRM, several works, such as Asefi et al. (2019) and Pinto and Stecca (2021), have studied the exploitation of OR for cost minimization in SRM management and WM. As this is a sector affected by many uncertainties, these are managed with different approaches. Gentile et al. (2023) consider the uncertainty in the demand using Robust Optimization theory for an operational problem. Caramia et al. (2023) extends the study considering the uncertainties that affect a more strategic problem, where a public authority has to redesign the WM network of an entire region. In our case, we deal with the uncertainties affecting the cost of serving a prospective new client, with an application to a WM network. This cost is evaluated by exploiting OR / ML methods and all relevant information in historical data.

Table 1 Contribution of methodological papers

References	Coll	Recycl	Uncert	OR	ML/AI	Cost
Yu et al. (2015)	X		X			X
Liu et al. (2022)	X		X			X
Verdonck et al. (2016)	X		X			X
Angelelli et al. (2022)	X			X		
Stecca and Kaihara (2021)	X	X		X		X
Mancini and Gansterer (2022)	X		X	X		X
Asefi et al. (2019)	X		X			
Nilashi et al. (2023)		X	X		X	
Bag et al. (2021)		X			X	
Ciardello et al. (2020)	X					X
Le et al. (2020)						X
Akkerman and Mes (2022)	X			X	X	
This paper	X	X	X	X	X	X

coll: collaboration, recycl: recycling, uncert: uncertainty, cost: cost allocation

The integration of AI/ML into DSS is a recent trend, as evidenced by studies such as Gupta et al. (2022), and its application is widespread in various sectors, including Waste Management, where AI/ML primarily serves as a predictive tool (Nilashi et al., 2023). AI/ML is used in conjunction with OR by Akkerman and Mes (2022) for estimating distribution cost in Traveling Salesman and Vehicle Routing Problems, by analyzing distance and demand features. This aligns with the observation that advanced tools like AI/ML are extremely useful in supporting the implementation of business strategies aimed at achieving Sustainable Development Goals (SDG), as discussed by Di Vaio et al. (2020) and shown in a case study by Cacciarelli and Boresta (2022).

To the best of our knowledge, despite the various significant applications of cost-to-serve analysis, no prior work has specifically addressed this particular problem in the manner our approach does. The framework developed in this study fills critical gaps identified in traditional methods of cost estimation, which frequently overlook the complete range of efficiency gains and client interdependencies. Through the integration of a broader spectrum of client features and the application of machine learning models, the framework can identify more complex relationships, resulting in more accurate estimations of clients' costs. This enhanced accuracy, coupled with a commitment to fairness, not only allows operational efficiency and cost-effectiveness for service providers but also fosters equitable treatment of clients.

The contribution given by the proposed framework to the cited literature can be summarized in the Table 1, where methodological papers are compared in their contributions. Our contribution covers several aspects, such as OR and ML/AI integrated methods, uncertainty in data, fair cost allocations, and collaboration in the logistics industry, in a framework for decision support.

The remainder of the paper is organized as follows: Sect. 2 describes the learning framework and its details; Sect. 3 introduces a real-case scenario application that features a novel logistic model presented in Appendix A; Sect. 4 reports the experimental results of applying the framework and provides insights and in-depth analysis of the client features importance; finally Sect. 5 gives some conclusions and research perspectives.

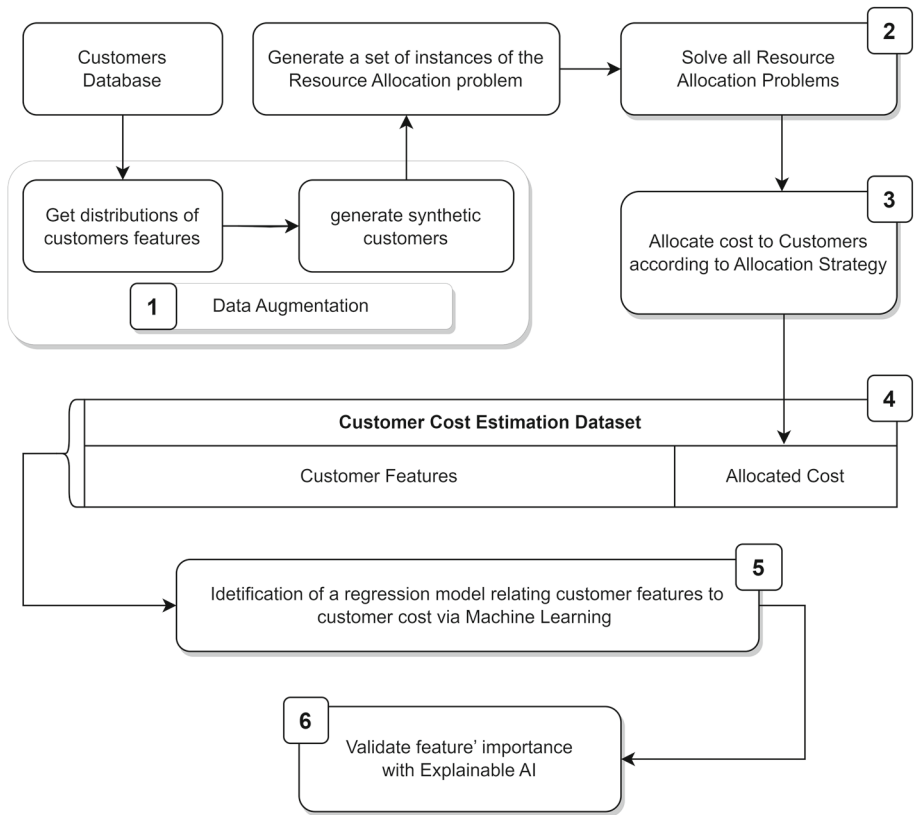


Fig. 1 OR-driven ML framework for client cost estimation

2 OR-driven ML framework for client cost estimation

The OR-driven ML Framework for Client Cost Estimation is a comprehensive approach that integrates OR models and ML techniques to enhance cost estimation processes in various industries.

This innovative framework aims to provide accurate cost estimates for new production or service requests from clients or prospects by leveraging historical data and taking into account the inherent variability in client features. The framework, illustrated in Fig. 1, consists of the following main steps:

1. Performing Data Augmentation
2. Generating and solving instances of the OR resource allocation problem
3. Selecting and using a fair Cost Allocation Strategy
4. Creating the ML Dataset
5. Training and Testing the ML model
6. Validating features importance with XAI techniques

These interconnected components are designed to work in synergy to achieve the aforementioned goals. Initially, the data provided by companies within the targeted industry is augmented with additional information that is potentially relevant, such as time windows

and service times, which may significantly impact costs. Data augmentation techniques are employed to create a more comprehensive dataset that encompasses a wider range of potential scenarios and factors that drive costs. By generating plausible values for missing or unrecorded features, the framework simulates an extensive variety of potential scenarios, resulting in a robust dataset for training the ML model. The generated feature values are based on reasonable probability distributions, ensuring they accurately reflect real-world conditions and constraints. Domain-specific knowledge and expert input are leveraged to refine these distributions to more closely match actual scenarios.

Following this, an OR model, tailored specifically to the industry problem at hand, is utilized to determine the optimal utilization of resources in serving clients. This model considers various constraints and objectives, such as minimizing operational costs and satisfying client time windows, to ensure the most efficient use of resources. A fair and reasonable cost allocation, derived from this optimization process, is then apportioned to each client, forming the target variable for the ML model.

The ML model, operating within the supervised learning paradigm, adopts a regression approach to forecast the cost of serving each client. It is trained on the enriched dataset, which includes a combination of existing client characteristics and new, instance-independent attributes. The choice of input features is deliberately selected based on the context and data available in the given industry, potentially including a broad spectrum of factors, from client demand patterns to operational time constraints, among others. Thus, the regression model is engineered to deduce the underlying cost function from the data, aiming to provide accurate, equitable, and scalable cost predictions for clients.

Explainable Artificial Intelligence (XAI) techniques are integrated into the framework to uncover the most critical factors affecting client costs. XAI aims to provide insights into the decision-making process of AI models, making their predictions more transparent and understandable for humans. By interpreting the trained ML model, the framework can identify the key features that drive cost allocation and reveal previously unknown relationships between client characteristics and costs, and these insights can be used to refine pricing strategies. For example, if XAI techniques reveal that time windows and service times are crucial factors affecting costs, the company may begin requesting this information from clients to produce better estimates.

The proposed framework addresses several shortcomings in traditional methods for cost estimation, such as the reliance on limited information like geographic distances or the magnitude of demand, which can lead to suboptimal and unfair resource allocation. By incorporating a wide range of client features and using ML models, the framework can capture complex relationships and interdependencies between clients. This results in more accurate and efficient cost estimates, that benefit both the service provider and the clients.

In practice, decision-makers typically rely on solving OR problems to determine optimal resource allocation among clients. However, the dynamic nature of many operational scenarios, makes it impractical to solve an OR problem for every new client, as the exact details of the upcoming schedule and how it will evolve are often unknown. Our framework overcomes this by learning from historical data and previously solved instances. This historical perspective makes the cost forecasting model more robust than estimates derived from considering a specific instance alone.

Another key aspect of this framework is that it allows companies to identify and understand the most impactful factors affecting client costs. Often, companies may not be aware of all the variables that influence the cost, relying only on readily available information, such as the client's distance from their facilities. This can result in suboptimal cost estimates that do not accurately reflect the true cost of providing services.

The proposed framework addresses this issue by incorporating a different set of client features, including both known and unknown variables, into the ML model, and by using XAI techniques to understand their impact on the costs.

In summary, the OR-driven ML Framework for Client Cost Estimation offers a powerful and flexible solution for companies seeking to improve their cost estimation processes. By leveraging historical data, generating missing information through data augmentation, and employing XAI techniques, this framework enables businesses to identify the most impacting factors for client costs and rely on more accurate, fair, and sustainable pricing strategies.

This approach benefits both the service provider and the clients, leading to improved resource allocation decisions and enhanced overall performance. The framework's versatility allows it to be adapted to various industries, making it a valuable tool for companies aiming to optimize their cost estimation processes and achieve a competitive edge in their respective markets.

3 Case study

To test the validity of the proposed framework, we consider a real case study involving an operator of a medium-sized waste company based in the province of Rome, Italy. Waste Management (WM) is a critical challenge concerning both the protection of the environment and the conservation of natural resources. For a comprehensive review of literature on strategic and tactical issues in solid WM, readers are referred to the surveys by Ghiani et al. (2014) and Das et al. (2019).

One major topic of interest within WM, besides Municipal Solid Waste Management (MSWM), is Industrial Waste Management (IWM), which encompasses tasks involved in collecting and transporting waste generated in industrial sites to sorting facilities where mixed materials are sorted to extract secondary raw materials. In this setting, waste companies typically serve their industrial clients according to a pull logic for waste container collection. A company truck picks up a client's waste container when the company's logistic service is contacted by the client for container pickup.

While planning and scheduling sorting facility operations represent the typical tactical task of IWM performed by OR models, the logistic cost of truck routing for waste pickup and delivery is instead the main operational cost of IWM that OR models need to optimize. In Sect. [Appendix A](#): , we present a new OR model that optimizes logistic operations within this IWM setting. In this scenario, the model finds the optimal set of routes for a fleet of trucks to pick up waste containers from a given set of industrial clients and deliver them to waste sorting facilities. This optimization enables the waste management company to minimize transportation costs, reduce environmental impact, and improve service quality for clients.

The problem discussed in Sect. [Appendix A](#): , which focuses on waste collection management, can also be considered as belonging to the class of Rollon-Rolloff Vehicle Routing Problems (RR-VRPs). The RR-VRP is studied in several papers, such as Bodin et al. (2000); Li et al. (2018) and Aringhieri et al. (2018), and involves tractors pulling large containers between client locations and disposal facilities. A feasible solution to this model simultaneously satisfies both the network system constraints and those associated with each waste-producing client and related disposal unit. Each graph node, corresponding either to a client or a waste sorting facility, has a time window bounding the time in which that node can be visited. On the other hand, truck units can haul either one or two waste containers, depending on each truck's capacity.

3.1 Dataset and data augmentation

The Waste Management firm provided access to data spanning from January 2020 to March 2022. The data comprises daily records of client service requests, including lists of clients served each day, their geographical information, locations of each client's waste sorting facility, and demand, expressed as the number of containers to be picked up. This dataset covers 668 operating days, with an average of 17.3 clients served, and 27 containers picked up each day.

In addition to the available information, this experiment investigates the impact of other variables on the resolution of the Vehicle Routing Problem (VRP) and consequently on the operating costs of the company. These variables, such as the visiting time windows for both the pickup and delivery nodes and their service times, have not been recorded in the company's database and are currently not considered by the contract management team when preparing commercial offers for clients.

However, it is evident that these variables, along with others not included in the dataset, could significantly affect the costs incurred by the business to serve a client. In fact, as highlighted in Malakar et al. (2020); Bacanin et al. (2021), feature selection and engineering are crucial for obtaining good results in ML training.

To incorporate this missing information into our analysis and enable the machine learning model to utilize it for cost prediction, we employed a data augmentation technique. Specifically, defining $J = \{j_1, j_2, \dots, j_n\}$ as the set of n potential extra features, we create κ copies of each operating day in our dataset and, for each copy, generate values for the j_{ih} missing value by random sampling from reasonable probability distributions. This approach allows us to simulate a range of possible scenarios for the same instance with different values for each extra feature, resulting in different costs for each scenario when solving the OR problems.

To further elucidate the data augmentation process used in our framework, we defined specific scenarios for the missing variables based on realistic and plausible situations in the context of waste management. For the time window of the pickup and delivery nodes (j_1 and j_2), we assume two distinct scenarios:

1. A long time window with a duration uniformly distributed in the range [6, 8] hours. The opening time for this scenario is uniformly distributed between 6:00 AM and 10:00 AM.
2. A short time window with a duration uniformly distributed in the range [2, 4] hours. The opening time for this scenario is uniformly distributed between 6:00 AM and 3:00 PM.

Regarding the service time for the pickup and delivery nodes (j_3 and j_4), we also assume two different situations:

1. A long service time scenario, where the service time is uniformly distributed in the range [30, 90] minutes.
2. A short service time scenario, where the service time is uniformly distributed in the range [10, 30] minutes.

It is challenging to estimate the distributions of missing data for generating their synthetic samples. These specific ranges and distributions were meticulously determined through consultation with domain experts, ensuring that they closely mirror real-world conditions in RL operations. For the data augmentation, we create $\kappa = 4$ copies of each operating day, with two copies sampled using the long scenarios for both the time window and service time, and two using the short scenarios. This process results in a dataset comprising 668×4 unique instances.

The data augmentation process, therefore, has a dual purpose. First, it allows testing the assumption that the extra features substantially impact the client cost. Second, it enables

the training of the machine learning model to incorporate this knowledge and provide more accurate cost predictions based on the newly generated data. It is essential to ensure that the values for the DA features are drawn from distributions that accurately reflect real-world conditions and constraints.

These distributions can be adjusted based on domain-specific knowledge or expert input to better adhere to realistic scenarios.

With these augmented scenarios forming a robust and varied dataset, we proceeded to the next step of creating the machine learning dataset, as described in the following subsection.

3.2 Creating the machine learning dataset

In this section, we describe the process of creating the dataset for the machine learning model, which involves determining target values and extracting relevant input features from both the original and augmented datasets.

3.2.1 Target values: cost allocation

We model the VRP problem with pickups, deliveries, and semi-soft time windows as outlined in Sect. [Appendix A](#): To determine how much of a route's total cost should be allocated to each of the clients served, we use the allocation strategy proposed and described in [Appendix B](#): . Specifically, we calculate the arithmetic average of three different allocation rules described therein: *Isolated Cost Allocation* (ICA), *Neighbors Savings* (NS), and *Normalized Marginal Allocation* (NMA). The resulting costs serve as the target values for training the machine learning model. The reader can refer to [Appendix B](#): for all further details.

3.2.2 ML model input features

The machine learning model's input features comprise instance-independent client characteristics, along with additional features engineered from the original dataset through DA. Focusing on instance-independent characteristics ensures that the model can generalize well to new clients and scenarios, rather than overfitting to the specific instances present in the training data.

The engineered features are constructed to facilitate the model's comprehension of the interplay between client attributes and the corresponding service costs. A key aspect of these features is their emphasis on the influence of client proximity in driving cost efficiencies. The feature columns are:

- *demand*: number of containers that the client asks to pick up
- *distHtoP*: shortest path distance in km from the depot (H) and the pickup node (P)
- *distHtoD*: shortest path distance in km from the depot (H) and the delivery node (D)
- *distPtoD*: shortest path distance in km from the pickup node and the associated delivery node
- *distSingleTour*: shortest path distance in km of the tour depot - pickup node - delivery node - depot
- *timeWindowPD*: sum of the widths of the pickup time window and delivery time window, expressed in minutes. For example, if the pickup node has a $9am - 3pm$ time window, and the delivery node has a $11am - 4pm$ time window, the variable *timeWindowPD* will have a value of $60 \cdot 6 + 60 \cdot 5 = 660$

- *serviceTimePD*: sum of the service time of the pickup node and delivery node, expressed in minutes
- *concentration20km*: number of other clients, across the whole dataset, that are within a 20km shortest path distance from the client
- *concentration50km*: number of other clients, across the whole dataset, that are within a 50km shortest path distance from the client
- *concentration100km*: number of other clients, across the whole dataset, that are within a 100km shortest path distance from the client

The 'concentration' columns in the dataset deserve special attention. These columns, indicating the number of other clients within specific distances, help the model understand cost allocation in relation to client proximity. The model's hypothesis is that clients who are geographically closer to each other can be served more efficiently, resulting in lower costs. This feature becomes particularly relevant when considering isolated clients who, due to their remoteness, may incur higher service charges.

The distance-related features, calculated using the Open Source Routing Machine (OSRM) (Luxen & Vetter, 2011), intuitively play a significant role in determining service costs.

To illustrate this, consider the scatter plot, as depicted in Fig. 2, shows the relationship between *distSingleTour* and *ServiceCost*. We can see that this relationship, while indicative, does not capture the full complexity of cost estimation. The plot clearly shows that distance alone is an incomplete predictor of service costs.

This limitation is evident in the upcoming numerical experiments detailed in Sect. 4. These experiments confirm that a broader range of features, beyond just distance, is necessary for

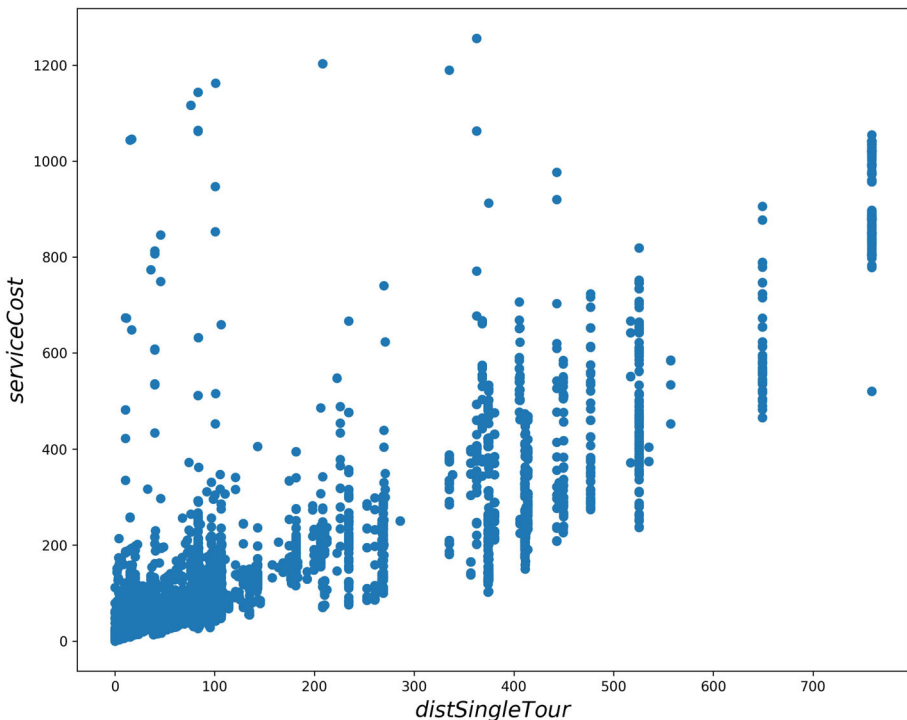


Fig. 2 Scatter plot *service cost–dist single tour*

accurate cost estimation. The ML model, by considering variables like client concentration, time windows, and service times, achieves a more accurate and comprehensive cost prediction.

4 Experimental results

This section presents the results from the application of our framework to the dataset detailed in Sects. 3 and 3.1. The objectives are to assess the framework's accuracy in service cost prediction and to identify key influencing variables using XAI techniques.

First, we explore the OR model implementation and our adopted strategy for cost allocation, forming the basis for determining service costs.

Following this, we focus on describing the training of our ML models commenting on their performance, and comparing them with two conventional methods typically used for cost estimation in this use-case context.

4.1 OR model and cost allocation strategy

We model the VRP problem with pickups, deliveries, and semi-soft time windows as described in [Appendix A](#): . Our analysis employs the dataset comprising both the original and the augmented data, following the augmentation methodology outlined in [3.1](#)

Model instances are coded in Python3 and tested on a PC running a 1.60GHz Intel Core-i5-10210U CPU with 16 GB RAM. For each of the 668 operating days, we solve different instances (one for each copy that differs in the values of the time windows and service time of each client) for a total of 2672 instances. These are solved via branch-and-cut using the Gurobi 9.5.2 solver hosted on a server running on an Intel Xeon Gold 6136 CPU @3.0 GHz with 250 GB RAM, with a solution time limit of 60 min.

When considering the model from a practical standpoint, where industry practitioners use it on a daily basis, the computational requirements of the model align reasonably well with the available time for resolution, being six hours in the considered use case. However, for larger instances, a natural trade-off exists between resolution time and the quality of the solution. Indeed, out of the total number of instances, 2172 (or 82% of the total) are solved with a gap of less than 0.05. Out of them, 1433 are exactly solved, finding the optimal solution.

In the following phase of the experiment, the ML model is trained only using the 2172 instances solved with the gap < 0.05 . While this approach might introduce a bias towards smaller, less complex instances, it leverages the most accurately solved examples to enhance model reliability. It would be interesting for future studies to explore how different trade-offs, such as accepting a higher gap and consequently a larger set of instances, might impact the model's performance.

The solutions of these instances, along with the objective function presented in [Appendix A](#): , allow for the determination of the total cost associated with a vehicle's route. Specifically, the total cost comprises the sum of all addends of the objective function that are pertinent to the particular vehicle. Notably, these cost elements are primarily associated with the route's total distance and duration, along with additional costs that may arise due to a delayed visit or the need for an overnight stay.

To determine how much of a route's total distance and duration should be allocated to each served client, we use the allocation strategy described in [Appendix B](#): . This allocation choice is just one of the possible allocation choices that can be made, and any other allocation

Table 2 ML models performance

Model	Metric	Training set	Test set
XGBoost	MAE	33.20	34.43
	MAPE	0.50	0.76
LightGBM	MAE	28.76	31.65
	MAPE	0.29	0.48

could be used without affecting the validity of the framework as long as it distributes costs fairly.

In the next section, we describe the training of the ML model and analyze the impact of each feature on the service cost prediction.

4.2 Model training and performance evaluation

For this study, we employed the XGBoost (Chen & Guestrin, 2016) and LightGBM (Ke et al., 2017) algorithms, both recognized for their exceptional performance on tabular data. These models frequently exhibit superior results when compared to traditional machine learning approaches (Shwartz-Ziv & Armon, 2022; Borisov et al., 2022).

For our dataset of 14,669 individual service records, we meticulously divided the data into training and testing sets using an 80–20 split. A critical measure was taken to ensure data integrity: if any record of a client on a specific day was placed in one set, all instances from that day were assigned consistently to prevent information leakage.

The implementation of both models was conducted using Python 3.8.10. The tuning of hyperparameters—such as *learning rate*, *number of estimators*, *subsample*, *max_depth*, *colsample_bytree*, and *alpha*—was accomplished through a random search, guided by five-fold cross-validation.

The random search spanned a pre-established array of values and ranges, allowing for comprehensive hyperparameter exploration. The optimal hyperparameter set was chosen based on its ability to minimize the Mean Absolute Error (MAE) across the validation folds.

The performance of the models is summarized in Table 2, where both MAE and Mean Absolute Percentage Error (MAPE) are reported for the training and test sets. Notably, the MAE exhibits close alignment between the training and test sets, underscoring the models' effective generalization capabilities. However, the MAPE demonstrates some variance, reflecting the sensitivity of this metric to the distribution of errors relative to actual service costs, as will be further discussed later in this section.

In addition to the advanced ML models, two baseline methodologies, herein referred to as *naive* and *HICE*, are employed to provide a comparative perspective on the performance of our framework.

The *naive* method estimates the client cost by exclusively considering a dedicated round trip route from the depot to the client's pickup and delivery locations, taking into account both the distance and the time required to complete the journey. This method also incorporates overnight costs for routes exceeding a single driving shift, the activation cost of the truck departing from the depot, and the time window enlargement cost. The time window enlargement is necessary when the time windows and service times of pickup and delivery nodes are incompatible with the travel time required to visit them. This approach is rather simple; it does not consider any savings that may result from serving another client in the neighborhood, nor does it take into account historical data.

Table 3 Performance comparison of ML and naive models. Bold stands for minimum values between each metric values

Model	MAE	MAPE
XGBoost	34.43	0.76
LightGBM	31.65	0.48
naive	192.24	3.80
HICE	92.64	0.94

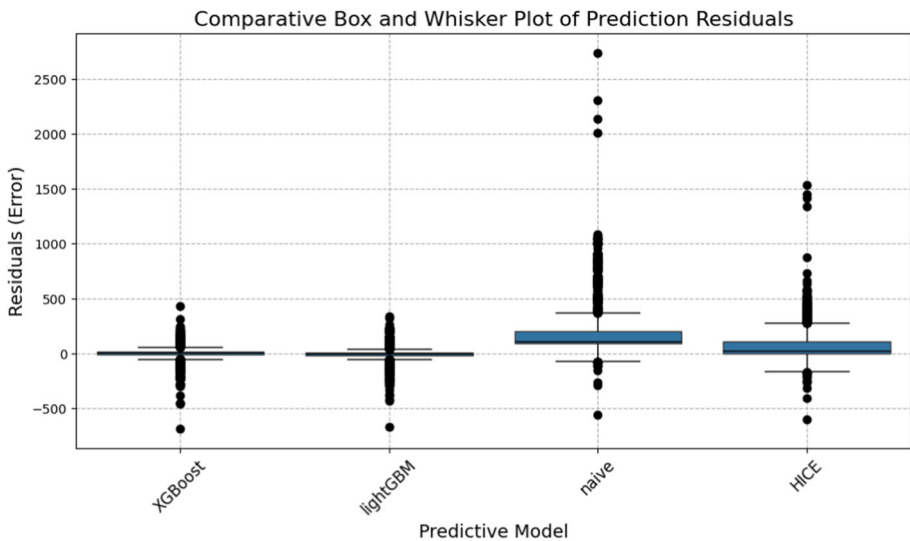


Fig. 3 Box and whisker plot of prediction residuals comparing XGBoost, LightGBM, naive and HICE

The Historical Integration Cost Estimation (HICE) method employs a data-driven strategy to estimate the client cost. This method involves selecting 10 random historical instances for each new client where he was not previously served. For each instance, a new instance that includes the additional client is generated. Then, by applying the optimization model outlined in Appendix A, the method recalculates the routes to identify the incremental costs associated with servicing the new client. This recalibration allows for an assessment of how each new client addition impacts operational costs, taking into account the variations inherent in historical routing patterns. The incremental costs discovered across these new instances are then averaged to provide an estimate of the service cost of the new client. While this method offers a more comprehensive analysis by factoring in variations in historical instances, it is computationally demanding, since it requires solving the routing optimization problem multiple times for each client.

Table 3 and Fig. 3 present the empirical results from these naive methods alongside the ML models. The performance metrics displayed in the table and the error distributions illustrated in the box and whisker plot delineate the accuracy and reliability of each method.

The data demonstrate that both machine learning (ML) models exhibit significantly lower absolute errors compared to the *naive* and *HICE* approaches, underscoring the sophistication and adaptability of ML in capturing the complexities of cost prediction in logistics services. Notably, the MAE, representing the actual monetary value of errors, is consis-

tently lower for ML models, highlighting their superior accuracy in monetary terms, which is particularly crucial in operational contexts.

To further substantiate our findings, we performed the Friedman (Friedman, 1937, 1940) and Quade (Quade, 1979) non-parametric statistical tests to determine whether the observed performance differences among the evaluated algorithms were statistically significant. These tests specifically aimed to evaluate the null hypothesis that there are no significant performance differences between the algorithms against the alternative hypothesis that at least one algorithm exhibits distinct performance.

The results from the Friedman test indicated highly significant differences among the algorithms, with chi-squared statistics significantly deviating from the null hypothesis. The Quade test corroborated these findings, revealing significant performance variations across the methods, with the largest p-value observed being 10^{-22} . This strongly suggests the presence of distinct performance differences among the evaluated models.

To further support these findings, a post-hoc analysis employing the Benjamini/Hochberg method was carried out. This analysis confirmed the significance of the observed differences, with all pairwise comparisons between the algorithms yielding p-values significantly below the 0.05 threshold. These statistical validations, implemented using the *statsmodels* (Seabold & Perktold, 2010) and *scikit_posthocs* (Terpilowski, 2019) Python libraries, affirm that the performance discrepancies between our machine learning models and the naive approaches are not only empirically observed but also statistically significant.

To conduct a more comprehensive analysis of the results, we examine the MAE and MAPE across various service cost ranges, recognizing that the distribution of service costs is heavily skewed towards the lower end.

Indeed, a significant proportion of client instances incur service costs below 100 units, which can disproportionately influence the MAPE metric. Table 4 provides a detailed breakdown of the MAE and MAPE for different service cost intervals, highlighting how the performance of the models varies across these ranges.

The analysis of Table 4 reveals valuable insights about the performance of the models across different service cost brackets. For all service cost ranges, both *naive* and *HICE* approaches exhibit markedly higher MAE and MAPE compared to the machine learning models. This indicates that the simplistic assumptions of *naive* and *HICE* are less effective, underscoring the superiority of the more sophisticated ML-based methodologies.

Moreover, *HICE* outperforms *naive* in all cost ranges. The latter is indeed an upper bound of *HICE* cost estimation, as *naive* strategy simplifies cost estimation by considering dedicated routes for service delivery, which inherently does not account for potential cost reductions through client aggregation. On the contrary, the *HICE* method leverages the OR model and its formulation to predict cost savings by simulating the incorporation of a new client into the historical routes, expecting to capitalize on the efficiencies of shared services.

4.2.1 Explainable AI

In response to the increasing emphasis on interpretability in machine learning, we utilized XAI techniques to demystify the predictive mechanisms of our models. Particularly, we utilized SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017), a powerful tool for interpreting model predictions. SHAP values, derived from cooperative game theory (Shapley, 1953), assign an importance value to each feature for a given prediction, offering insights into how each feature influences the model's output.

Figure 4 presents a summary plot for our LightGBM model, combining feature importance with their effects on the predicted service costs.

Table 4 Breakdown of MAE and MAPE metrics

Service cost range	Test records	Model	MAE	MAPE
0–100	1687 (58%)	XGBoost	19,76	1,15
		LightGBM	15,08	0,68
		naive	110,24	5,76
		HICE	26,76	1,1
101–200	420 (15%)	XGBoost	48,61	0,34
		LightGBM	50,16	0,35
		naive	214,09	1,45
		HICE	155,72	1,04
201–500	580 (20%)	XGBoost	46,58	0,15
		LightGBM	44,67	0,15
		naive	252,39	0,90
		HICE	167,13	0,6
501–1062	207 (7%)	XGBoost	90,37	0,11
		LightGBM	91,30	0,11
		naive	650,10	0,75
		HICE	293,72	0,35

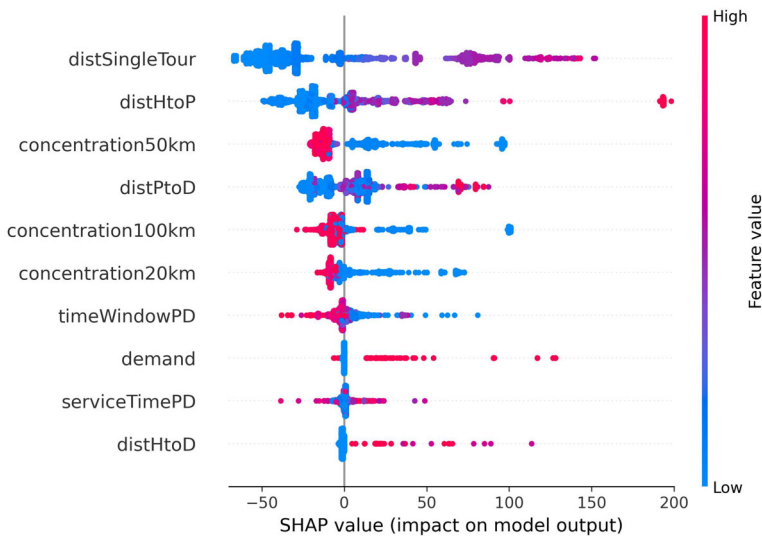


Fig. 4 SHAP values explaining feature importance in the LightGBM model

The plot reveals that the feature *distSingleTour*, indicating the shortest complete tour distance, is the most impactful. Lower values of *distSingleTour* correlate with negative SHAP values, while higher values correspond to positive SHAP values, suggesting that shorter tour distances typically lead to lower service costs.

Client density features, namely *concentration20km*, *concentration50km*, and *concentration100km*, display similar patterns. They underscore the model’s recognition of logistical efficiencies gained through servicing clustered clients. Higher client concentrations often

result in shared service routes and thus lower predicted costs, reflecting the hypothesized advantage of geographical client proximity on service cost efficiency.

Time-related features like *timeWindowPD* and *serviceTimePD*, though less influential, still reflect the nuances of scheduling constraints and service durations on cost predictions.

It is important to note that a similar SHAP plot analysis for the XGBoost model exhibited analogous behavior in terms of feature impacts. However, for brevity, we present only the LightGBM model's plot here. This consistency across models reinforces our confidence in the robustness and generalizability of the observed feature importance trends.

5 Conclusions

In conclusion, this study represents a significant advancement in integrating Operations Research and Machine Learning to enhance decision-making processes in the realm of sustainable business management, especially within the logistics and service sectors. This integration effectively combines the operational efficiency of OR with the predictive power of ML in service cost estimation, considering the diverse characteristics of different clients. This innovative approach is instrumental for businesses operating in the dynamic and complex environment of a circular economy.

5.1 Theoretical and practical implications

Theoretically, our research introduces a novel framework that integrates OR and ML/AI methodologies. This framework addresses several key challenges, such as data uncertainty and equitable cost allocation, prevalent in various industry contexts. Practically, it serves as a versatile decision-support tool that enables rapid and precise cost estimation, enhancing margin assessments and facilitating more informed client-specific cost allocation. Consequently, it provides a practical solution to the challenges of low margins and uncertainties, which are often encountered in circular economy processes across different service sectors.

5.2 Research contributions

Our study presents a unique integration of OR and ML, significantly improving service cost forecasting in the circular economy context. This integration not only boosts the accuracy of cost estimations but also brings a nuanced understanding of how different client characteristics influence these costs. Additionally, the incorporation of XAI methods in our framework increases the transparency and interpretability of the ML models. This dual approach is a substantial contribution to both academic research, by showcasing an innovative use of integrated methodologies, and to industry practices, by offering a sophisticated yet practical approach to cost management in various service sectors.

5.3 Practical advantages

The framework developed in this study addresses some critical gaps found in traditional methods of cost estimation, which often fail to consider the full spectrum of efficiency gains and client interdependencies. By integrating a wider range of client features and employing ML models, the framework can discern more complex relationships and interactions, leading

to more accurate client cost estimations. This enhanced accuracy, in conjunction with the principles of fairness, not only improves operational efficiency and cost-effectiveness for service providers but also promotes equitable treatment of clients.

5.4 Research limitations and future directions

This study, while providing valuable insights, is subject to several certain that also pave the way for future research directions. A primary theoretical limitation stems from our reliance on a simplified cost allocation rule. Due to the practical challenges in applying more complex approaches like Shapley Values, which demand extensive computational resources, we adopted a more straightforward rule. While effective in our context, its optimality and general applicability across various scenarios remain unverified. Future research could focus on exploring alternative strategies for cost allocation, including high-quality approximations to Shapley values, or further validation of our current approach.

Practically, the current application of our framework is confined to the waste management industry. This limitation highlights the need for future studies to expand its applicability to a diverse range of service industries, thereby enhancing its adaptability to different operational contexts. Additionally, the efficacy of our framework is highly dependent on the availability and quality of historical data, which is a critical factor for the effective training of the machine learning model. This dependency underscores a limitation, as the framework may not apply to companies without established data collection practices.

Furthermore, conducting sensitivity analyses on feature importance in cost allocation strategies could provide deeper insights into the effectiveness of different methodologies.

In summary, this study lays a robust foundation for future research and practical applications in diverse logistics and service industries, offering innovative and sustainable solutions in service cost forecasting and allocation. It demonstrates the potential of integrating OR and ML as a decision-support tool, contributing significantly to the evolution of sustainable business practices and the advancement towards a circular economy.

Appendix A: Operational problem definition and modeling

This appendix provides a detailed description of a novel mixed-integer linear programming model designed for optimizing route planning in pick-up and delivery services. This general problem is one of the main operational tasks of supply chain management and is known as the Vehicle Routing Problem (VRP), a NP-Hard and well-known combinatorial optimization problem (Toth & Vigo, 1998). The presented formulation is intended to find the optimal set of routes for a single depot fleet of trucks to pick up goods from a given set of customers and deliver them to another given set of delivery nodes. A feasible solution of this model simultaneously satisfies the network system constraints, transport capacity constraints, and constraints associated with each customer and delivery location. Each graph node, representing either a customer or a delivery node, has a time window that constrains the timeframe during which the node can be visited. Truck units are heterogeneous in terms of operating and emissions costs and can haul goods according to specific hauling capacities. Non-linear costs are factored in for truck drivers, using a two-part tariff comprised of a fixed fee plus a cost per driving hour. Moreover, an additional overnight cost can be incurred, potentially allowing trucks to take multiple driving shifts to reach remote locations. If a node's time window is particularly small, potentially causing infeasibility or limiting route savings, the window can

be extended with an additional cost per hour. The presented formulation can be regarded as a variant of the VRP with pickups and deliveries (VRPPD) (Toth & Vigo, 2002) which only considers penalties on late arrivals while allowing waiting for early arrivals without any cost. This specific problem is known as the Vehicle Routing and Scheduling Problem with Semi Soft Time Windows (VRPSSTW) (Qureshi et al., 2010). A logistic service request i is characterized by two nodes, node i and node $n + i$, corresponding, respectively, to the pickup and delivery stops of the request. Different nodes may represent the same geographical location. The set of pickup nodes is denoted by $P = \{1, \dots, n\}$ and the set of delivery nodes is denoted by $D = \{n + 1, \dots, 2n\}$. If request i consists of transporting d_i units from i to $n + i$, let $l_i = d_i$ and $l_{n+i} = -d_i$. Furthermore, let K represents the set of vehicles. The objective of the VRP is to determine routes serving all the customers, respecting vehicle and user constraints, while minimizing the total travel costs.

To provide a clearer introduction to the formulation and its additional features compared to those presented in Toth and Vigo (2002) and Qureshi et al. (2010), the model notation for parameters, indexes, and variables is outlined below.

- n : number of customers to be served
- $P = \{1, \dots, n\}$: set of pickup nodes
- $D = \{n + 1, \dots, 2n\}$: set of delivery nodes
- $N = P \cup D$: overall set of pickup and delivery nodes to visit
- o : depot as a departing node
- d : depot as a returning node
- $od = o \cup d$
- $No = N \cup o$
- $Nd = N \cup d$
- $Nod = o \cup N \cup d$: overall set of graph nodes
- K : set of heterogeneous vehicles
- $[a_i, b_i]$: service time window for node i
- $c_{i,j}$: cost of arch i, j as the sum of distance and duration costs
- $t_{i,j}$: arch i, j travel time
- s_i : service time at node i
- l_i : pickup or delivery demand at node i
- C_k : carrying capacity of vehicle k
- u_k : activation cost of vehicle k related to expected operating and emissions costs
- O : overnight cost
- R : maximum number of driving hours in a shift
- \mathcal{Z} : maximum number of subsequent driving shifts
- Γ : time window enlargement cost per hour

The model consider the following variables.

- $x_{i,j,k} \in \{0, 1\}$: equal to 1 if vehicle k crosses arch (i, j) , 0 otherwise
- $T_{i,k} \in \mathbb{R}^+$: service starting time at node i by vehicle k
- $L_{i,k} \in \mathbb{Z}^+$: load of vehicle k when leaving node i
- $y_k \in \mathbb{R}^+$: bounds the time span between departure and returning to depot of vehicle k
- $z_k \in \mathbb{Z}^+$: number of subsequent driving shifts of vehicle k
- $o_k \in \mathbb{Z}^+$: number of subsequent overnights stay for vehicle k
- $\gamma_i \in \mathbb{R}^+$ node i time window upper bound enlargement

The model minimizes the sum of several transport costs and is detailed as following:

$$\min Z = \sum_{k \in K} \sum_{i \in \text{Nod}} \sum_{j \in \text{Nod}} c_{i,j} x_{i,j,k} + \sum_{k \in K} (u_k z_k + O o_k) + \sum_{k \in K} y_k + \Gamma \sum_{i \in N} \gamma_i \quad (\text{A.1})$$

s.t.

$$\sum_{i \in \text{od}} L_{i,k} = 0 \quad \forall k \in K \quad (\text{A.2})$$

$$\sum_{k \in K} \sum_{j \in \text{Nd}} x_{i,j,k} = 1 \quad \forall i \in P \quad (\text{A.3})$$

$$\sum_{j \in N} x_{i,j,k} = \sum_{j \in N} x_{j,n+i,k} \quad \forall k \in K, i \in P \quad (\text{A.4})$$

$$\sum_{j \in P} x_{0,j,k} \leq 1 \quad \forall k \in K \quad (\text{A.5})$$

$$\sum_{i \in D} x_{i,2n+1,k} = \sum_{j \in P} x_{0,j,k} \quad \forall k \in K \quad (\text{A.6})$$

$$\sum_{i \in \text{No}} x_{i,j,k} = \sum_{i \in \text{Nd}} x_{j,i,k} \quad \forall k \in K, j \in N \quad (\text{A.7})$$

$$T_{i,k} + s_i + t_{i,j} - T_{j,k} \leq M(1 - x_{i,j,k}) \quad \forall k \in K, i \in \text{Nod}, j \in \text{Nod}, i \neq j \quad (\text{A.8})$$

$$a_i \leq T_{i,k} \leq b_i - s_i + \gamma_i \quad \forall k \in K, i \in N \quad (\text{A.9})$$

$$T_{i,k} + s_i + t_{i,n+i} \leq T_{n+i,k} \quad \forall k \in K, i \in P \quad (\text{A.10})$$

$$L_{i,k} + l_j - L_{j,k} \leq M(1 - x_{i,j,k}) \quad \forall k \in K, i \in N, j \in N, i \neq j \quad (\text{A.11})$$

$$l_i \sum_{j \in N, j \neq i} x_{i,j,k} \leq L_{i,k} \leq C_k \quad \forall k \in K, i \in P \quad (\text{A.12})$$

$$L_{i,k} \leq (C_k + l_i) \sum_{j \in N, j \neq i} x_{i,j,k} \quad \forall k \in K, i \in D \quad (\text{A.13})$$

$$T_{|N|+1,k} - T_{0,k} \leq y_k \quad \forall k \in K \quad (\text{A.14})$$

$$\sum_{i \in N, i \neq j} \sum_{j \in N, j \neq i} (s_i + t_{i,j}) x_{i,j,k} \leq R z_k \quad \forall k \in K \quad (\text{A.15})$$

$$z_k \leq Z \left(\sum_{j \in P} x_{0,j,k} \right) \quad \forall k \in K \quad (\text{A.16})$$

$$z_k \geq \sum_{j \in P} x_{0,j,k} \quad \forall k \in K \quad (\text{A.17})$$

$$o_k \geq 0 \quad \forall k \in K \quad (\text{A.18})$$

$$o_k \geq z_k - 1 \quad \forall k \in K \quad (\text{A.19})$$

$$\gamma_i \geq 0 \quad \forall i \in N \quad (\text{A.20})$$

$$x_{0,j,k} = 0 \quad \forall k \in K, j \in D \quad (\text{A.21})$$

$$x_{i,i,k} = 0 \quad \forall i \in \text{Nod}, k \in K \quad (\text{A.22})$$

$$x_{i,0,k} = 0 \quad \forall i \in N, k \in K \quad (\text{A.23})$$

The objective function (A.1) aims to minimize the sum of routing costs, which include emissions and operating costs per kilometer and hour, driving shift activation and overnight

costs, and time window enlargement costs. Additionally, the function encompasses the unit cost of the auxiliary variable y , which bounds the time span between a vehicle's departure from and return to the depot for each utilized vehicle.

Constraints (A.2) ensure that each truck is empty when exiting or entering the depot; constraints (A.3) guarantee that each customer $i \in P$ is served by one and only one truck; constraints (A.4) describe the network flow conservation and simultaneously ensure that each pickup is delivered to its respective delivery location; constraints (A.5) guarantee that each used truck departs from the depot to reach a pickup node; constraints (A.6) force used vehicles to return to the depot; constraints (A.7) ensure network flow conservation; for all subsequently visited nodes, constraints (A.8) introduce time coherence between arrivals, service and travel times; constraints (A.9) mandate that each node i is visited within its time window while considering service time and the option of extending the upper bound b_i ; constraints (A.10) ensure that delivery occurs later than pickup; constraints (A.11) introduce truck load coherence between each subsequently visited node; constraints (A.12) and (A.13) limit truck load capacity and guarantee truck load coherence when visiting a pickup and a delivery node respectively; constraints (A.14) bound vehicle time spans between the exit and the return to depot; constraints (A.15) limit each vehicle travel time with respect to the driving shifts and their maximum driving hours; constraints (A.16) bound the maximum number of subsequent shifts for each used truck; constraints (A.17) ensure the activation of at least one driving shifts of each used truck; constraints (A.18) and (A.19) imply that for each truck, the number of overnights is zero unless it is used for more than one driving shift; constraints (A.20) ensure that time windows can only be extended; constraints (A.21) prevent trucks from traveling directly from the depot to a delivery node; constraints (A.22) avoid node loops, and constraints (A.23) ensure that no vehicle returns to the depot as a departing node.

The reader can refer to the reviews presented in Braekers et al. (2016) and Tan and Yeh (2021) to find relevant literature, recent trends and solution methodologies in the field of VRPs and some well-known variants.

Appendix B: Fair cost allocation rule

Given the solution of a resource allocation problem, it is important to define a rule that allocates a fair portion of the overall production or service cost to each customer. The problem in transportation logistics is determining how much of a route's total cost should be allocated to each customer served.

This cost allocation problem is very challenging and has been addressed in past and recent literature using different strategies, ranging from relatively simple to state-of-the-art approaches (Kilby, 2018).

Such cost allocation problems typically exhibit two features: (i) costs must be allocated exactly, with no profit or deficit; (ii) there is no objective basis at hand for attributing costs directly to specific products or services. The goal is to devise criteria and methods for solving these problems in a just, equitable, fair, and reasonable manner (Young, 1985). Cost allocation is thus ultimately concerned with fairness. The methods and principles of cost allocation that are likely to find acceptance must somehow be grounded in primitive, common-sense ideas of fairness and equity. In the book *Cost allocation: methods, principles, applications* by Young (1985), the authors propose several methods to address this widespread problem in different scenarios. However, they state that whether the reader finds the proposed cost allocation principles compelling will depend to some degree on taste, for fairness is partly in the eye of

the beholder. Indeed, although there are many formal definitions of fairness and solutions to such cost allocation problems, there is no absolute definition, as fairness is always a context-related concept. That implies that no universal metric will indicate that one solution is more fair than another.

Considering the logistic service scenario, these problems are usually cast in the form of a cooperative game theory model (Ciardiello et al., 2020; Le et al., 2020). According to this approach, an ideal solution is well-defined and is based on the Shapley value (Shapley, 1953) of the Traveling Salesman Game (TSG) (Potters et al., 1992). Essentially, the Shapley value of player/customer i is an average of all marginal costs attributable to player i across all possible coalitions that include player i . The Shapley value is the only attribution method that satisfies the properties of *efficiency*, *symmetry*, *dummy*, and *additivity* (Shapley, 1953), which together are considered a definition of a fair allocation, making it a preferred choice for a fair cost allocation scheme (Popescu & Kilby, 2020). Unfortunately, this method has exponential computational complexity, rendering it practical only for small-scale examples involving a few customers. In real-case applications, scenarios involving several dozen or even a few hundred customers are more typical. This has led to substantial research efforts dedicated to finding approximations with lower computational complexity. In Castro et al. (2009) an approximation approach to Shapley Values through Monte Carlo methods is presented, while novel methodologies to generate high-quality approximations to the Shapley values are presented in Popescu and Kilby (2020) and Levinger et al. (2021).

Given the scope of this paper, and the computational burden of computing the Shapley values, the authors of this paper developed a cost allocation strategy outside the game theory paradigm. The strategy presented in the following is computationally tractable and appears to be fair as it is easy to prove that it satisfies three of the four desirable properties of the Shapley value approach (Shapley, 1953), such as *efficiency*, *symmetry*, and *null player* property. Indeed, regardless of any assumption or setting, satisfying these desirable properties is sufficient and reasonable to present the following cost allocation approach as fair.

The proposed approach applies to the solution of a Vehicle Routing Problem (VRP) with Pickup and Delivery, where a customer $i \in \{1, \dots, n\}$ is identified by two distinct nodes: the pickup node P_i and the delivery node D_{n+i} .

The proposed allocation strategy employs a weighted average of the outcomes from three different allocation rules, which we name as follows:

- *Isolated Cost Allocation* (ICA)
- *Neighbors Savings* (NS)
- *Normalized Marginal Allocation* (NMA)

The solution of the VRP can generate multiple routes to pass through all customer locations. The allocation strategy is then applied to each of these routes assigned to a specific vehicle k . To better introduce the allocation rules, consider the following notation:

- n = number of customer to be served
- $P = \{P_1, \dots, P_n\}$: set of pickup nodes
- $D = \{D_{n+1}, \dots, D_{2n}\}$: set of delivery nodes
- $N = P \cup D$: overall set of pickup and delivery nodes to visit
- o : depot as a departing node
- d : depot as a returning node
- $Nod = o \cup N \cup d$: overall set of graph nodes
- $c_{i,j}$: cost of arch i, j as the sum of distance and duration costs

The total cost (i.e. C_{tot}) of the solution is defined as follows.

$$C_{tot} = \sum_{i \in \text{Nod}} \sum_{j \in \text{Nod}} c_{i,j} x_{i,j}$$

Where $x_{i,j} \in \{0, 1\}$ is equal to 1 if vehicle k crosses arch (i, j) , 0 otherwise. Considering this notation, the three allocation rules are presented below.

The *Isolated Cost Allocation* (ICA) rule assigns costs to each customer i in proportion to the shortest path distances of the route that starts from the depot, moves to the pickup node P_i , then to the delivery node D_{n+i} , and finally returns to the depot. Considering a customer i , we define:

$$C_i^{ICA} = C_{o,P_i} + C_{P_i,D_{n+i}} + C_{D_{n+i},d}$$

Once C_i^{ICA} is computed $\forall i \in \{1, \dots, n\}$, these values are normalized to get the *Isolated Cost Allocation* portion of C_{tot} to allocate to each customer i , namely $\%C_{iICA}$

$$\%C_{iICA} = \frac{C_i^{ICA}}{\sum_{i \in \{0, \dots, n\}} C_i^{ICA}}$$

Therefore, *Isolated Cost Allocation* rule assigns a cost C_{iICA} to customer i such that:

$$C_{iICA} = \%C_{iICA} * C_{tot}$$

This rule can serve as an initial proxy to allocate a proportion of the total route cost. This rule assigns a cost that is proportionate to both the P_i and D_{n+i} distances and takes into account these nodes as independent of the other visited ones, embedding the non-cooperative scenario. At the same time, this represents its main drawback, as the *ICA* rule does not consider the reduction in costs due to proximity to other served customers, both in terms of pickup and delivery.

Neighbors Savings rule is intended to recover the main drawback of the *ICA* rule. This is achieved by evaluating the marginal cost of visiting P_i and D_{n+i} in the vehicle tour. With this objective, two quantities are defined for each customer i , namely $C_{deviation}$ and C_{link} . $C_{deviation}$ is the sum of all travel distances to visit both P_i and D_{n+i} as deviations from the tour of k not considering customer i . This is the sum of all distances associated to the graph edges crossed by k in its tour that have P_i or D_{n+i} as either departing or arriving node. Accordingly, $C_{deviation}$ is defined $\forall i \in \{1, \dots, n\}$ as follows:

$$C_{deviation} = \sum_{l \in \text{Nod}} C_{l,P_i} x_{l,P_i} + \frac{\sum_{l \in \text{Nod}} C_{P_i,l} x_{P_i,l} + \sum_{l \in \text{Nod}} C_{l,D_{n+i}} x_{l,D_{n+i}}}{1 + x_{P_i,D_{n+i}}} + \sum_{l \in \text{Nod}} C_{D_{n+i},l} x_{D_{n+i},l}$$

C_{link} is the sum of all distances associated to the graph edges that k should cross to complete its tour without visiting P_i and D_{n+i} while respecting the nodes visiting order of the original solution. Accordingly, C_{link} is defined $\forall i \in \{1, \dots, n\}$ as follows:

$$C_{link} = \left[\sum_{l \in \text{Nod}} \sum_{h \in \text{Nod}} C_{l,h} (x_{l,P_i} x_{P_i,h}) + \sum_{l \in \text{Nod}} \sum_{h \in \text{Nod}} C_{l,h} (x_{l,D_{n+i}} x_{D_{n+i},h}) \right] (1 - x_{P_i,D_{n+i}}) + \left[\sum_{l \in \text{Nod}} \sum_{h \in \text{Nod}} C_{l,h} (x_{l,P_i} x_{D_{n+i},h}) \right] (x_{P_i,D_{n+i}})$$

Given the above definitions, we define C_i^{NS} as the marginal distance contribution of i :

$$C_i^{NS} = C_{tot} - C_{deviation} + C_{link}$$

Once C_i^{NS} is computed $\forall i \in \{1, \dots, n\}$, these values are normalized to get the *Neighbors Savings* portion of C_{tot} to allocate to each customer i :

$$\%C_{iNS} = \frac{C_i^{NS}}{\sum_{i \in \{1, \dots, n\}} C_i^{NS}}$$

Therefore, *Neighbors Savings* rule assigns a cost C_{iNS} to customer i such that:

$$C_{iNS} = \%C_{iNS} * C_{tot}$$

The main drawback of the *NS* rule is the risk of not being fair to customers in proximity of the depot that are involved in a tour that reaches distant locations. *NS* would assign a portion of cost that is related to the value of C_{tot} which is expected to be high in such a tour.

Normalized Marginal Allocation rule is intended to recover the mentioned drawback of the *NS* rule. This is accomplished by computing the marginal distance contribution of i without considering C_{tot} . Therefore, *NMA* employs the *NS* rule except for considering $C_i^{NMA} = \max(0, C_{deviation} - C_{link})$. Therefore, once C_i^{NMA} is computed $\forall i \in \{0, \dots, n\}$, these values are normalized to get the *Normalized Marginal Allocation* portion of C_{tot} to allocate to each customer i :

$$\%C_{iNMA} = \frac{C_i^{NMA}}{\sum_{i \in \{0, \dots, n\}} C_i^{NMA}}$$

Hence, *Normalized Marginal Allocation* rule assigns a cost C_{iNMA} to customer i such that:

$$C_{iNMA} = \%C_{iNMA} * C_{tot}$$

The final output of the proposed strategy is a cost allocation proportion that is a weighted average of the outcomes of the *ICA*, *NS*, and *NMA* rules. Numerical experiments have analyzed the effects of different weights to address the fairness of this allocation technique. Intuitively, equal weights provide a fair balance between the objectives and drawbacks of all rules.

The final average follows a pattern similar to the *Isolated Cost Allocation* rule, indicating that the distance between the depot and the customers is the parameter with the greatest weight. But this is dampened by the other two rules, succeeding in the goal of finding a consistent and easy-to-compute cost allocation strategy that is outside the game theory paradigm.

Acknowledgements This work has been partially supported by EU POR FESR program of LAZIO Region on “Research Groups 2020” through the project PIPER - Piattaforma intelligente per l’ottimizzazione di operazioni di riciclo [Grant No. A0375-2020-36611, CUP B85F21001480002] and partially supported by PNRR MUR project PE0000013-FAIR.

Funding Open access funding provided by Consiglio Nazionale Delle Ricerche (CNR) within the CRUI-CARE Agreement. Grant no. (A0375-2020-36611, CUP B85F21001480002), Ministero dell’Istruzione, dell’Università e della Ricerca (PE0000013-FAIR).

Data availability Data used in this work are available by request to authors.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akkerman, F., & Mes, M. (2022). Distance approximation to support customer selection in vehicle routing problems. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-022-04674-8>
- Angelelli, E., Morandi, V., & Speranza, M. G. (2022). Optimization models for fair horizontal collaboration in demand-responsive transportation. *Transportation Research Part C: Emerging Technologies*, *140*, 103725.
- Aringhieri, R., Bruglieri, M., Malucelli, F., & Nonato, M. (2018). A special vehicle routing problem arising in the optimization of waste disposal: A real case. *Transportation Science*, *52*, 277–299.
- Asefi, H., Lim, S., Maghrebi, M., & Shahparvari, S. (2019). Mathematical modelling and heuristic approaches to the location-routing problem of a cost-effective integrated solid waste management. *Annals of Operations Research*, *273*, 75–110.
- Bacanin, N., Stoean, R., Zivkovic, M., Petrovic, A., Rashid, T. A., & Bezdán, T. (2021). Performance of a novel chaotic firefly algorithm with enhanced exploration for tackling global optimization problems: Application for dropout regularization. *Mathematics*, *9*, 2705.
- Bag, S., Pretorius, J. H. C., Gupta, S., & Dwivedi, Y. K. (2021). Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. *Technological Forecasting and Social Change*, *163*, 120420.
- Barros, M. V., Salvador, R., do Prado, G. F., de Francisco, A. C., & Piekarski, C. M. (2021). Circular economy as a driver to sustainable businesses. *Cleaner Environmental Systems*, *2*, 100006.
- Bodin, L., Mingozzi, A., Baldacci, R., & Ball, M. (2000). The rollon-rolloff vehicle routing problem. *Transportation Science*, *34*, 271–288.
- Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2022). Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2022.3229161>
- Braekers, K., Ramaekers, K., & Van Nieuwenhuysse, I. (2016). The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering*, *99*, 300–313.
- Cacciarelli, D., & Boresta, M. (2022). What drives a donor? A machine learning-based approach for predicting responses of nonprofit direct marketing campaigns. *Journal of Philanthropy and Marketing*, *27*, e1724.
- Caramia, M., Pinto, D. M., Pizzari, E., & Stecca, G. (2023). Clustering and routing in waste management: A two-stage optimisation approach. *EURO Journal on Transportation and Logistics*, *12*, 100114.
- Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research*, *36*, 1726–1730.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining (pp. 785–794).
- Ciardello, F., Genovese, A., & Simpson, A. (2020). A unified cooperative model for environmental costs in supply chains: The Shapley value for the linear case. *Annals of Operations Research*, *290*, 421–437.
- Das, S., Lee, S. H., Kumar, P., Kim, K. H., Lee, S. S., & Bhattacharya, S. S. (2019). Solid waste management: Scope and the challenge of sustainability. *Journal of Cleaner Production*, *228*, 658–678.
- Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, *121*, 283–314.

- Fatimah, Y. A., Govindan, K., Murniningsih, R., & Setiawan, A. (2020). Industry 4.0 based sustainable circular economy approach for smart waste management system to achieve sustainable development goals: A case study of Indonesia. *Journal of Cleaner Production*, 269, 122263.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32, 675–701.
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, 11, 86–92.
- Gentile, C., Pinto, D. M., & Stecca, G. (2023). Price of robustness optimization through demand forecasting with an application to waste management. *Soft Computing*, 27, 13013–13024.
- Ghiani, G., Laganà, D., Manni, E., Musmanno, R., & Vigo, D. (2014). Operations research in solid waste management: A survey of strategic and tactical issues. *Computers & Operations Research*, 44, 22–32.
- Gupta, S., Modgil, S., Bhattacharyya, S., & Bose, I. (2022). Artificial intelligence for decision support systems in the field of operations research: Review and future scope of research. *Annals of Operations Research*, 308, 1–60.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 3146–3154.
- Kilby, Phil, & Popescu, D. (2018). *Efficient computation of cost allocations for the vehicle routing problem* (p. 15). Melbourne: AMSI Optimise.
- Lai, K. H., & Wong, C. W. (2012). Green logistics management and performance: Some empirical evidence from Chinese manufacturing exporters. *Omega*, 40, 267–282.
- Le, P. H., Nguyen, T. D., & Bektaş, T. (2020). Efficient computation of the Shapley value for large-scale linear production games. *Annals of Operations Research*, 287, 761–781.
- Levinger, C., Hazon, N., & Azaria, A. (2021). Efficient computation and estimation of the Shapley value for traveling salesman games. *IEEE Access*, 9, 129119–129129.
- Li, H., Jian, X., Chang, X., & Lu, Y. (2018). The generalized rollon-rolloff vehicle routing problem and savings-based algorithm. *Transportation Research Part B: Methodological*, 113, 1–23.
- Liu, H., Yu, Y., Benjaafar, S., & Wang, H. (2022). Price-directed cost sharing and demand allocation among service providers with multiple demand sources and multiple facilities. *Manufacturing & Service Operations Management*, 24, 647–663.
- Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (vol. 30).
- Luxen, D., & Vetter, C. (2011). Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*, ACM, New York, NY, USA (pp. 513–516). <https://doi.org/10.1145/2093973.2094062>
- Malakar, S., Ghosh, M., Bhowmik, S., Sarkar, R., & Nasipuri, M. (2020). A GA based hierarchical feature selection approach for handwritten word recognition. *Neural Computing and Applications*, 32, 2533–2552.
- Mancini, S., & Gansterer, M. (2022). Bundle generation for last-mile delivery with occasional drivers. *Omega*, 108, 102582.
- Milios, L. (2018). Advancing to a circular economy: Three essential ingredients for a comprehensive policy mix. *Sustainability Science*, 13, 861–878.
- Nilashi, M., Baabdullah, A. M., Abumalloh, R. A., Ooi, K. B., Tan, G. W. H., Giannakis, M., & Dwivedi, Y. K. (2023). How can big data and predictive analytics impact the performance and competitive advantage of the food waste and recycling industry? *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05272-y>
- Pinto, D. M., & Stecca, G. (2021). Optimal planning of waste sorting operations through mixed integer linear programming. *Graphs and Combinatorial Optimization: From Theory to Applications: CTW2020 Proceedings*. https://doi.org/10.1007/978-3-030-63072-0_24
- Popescu, D. C., & Kilby, P. (2020). Approximation of the Shapley value for the Euclidean travelling salesman game. *Annals of Operations Research*, 289, 341–362.
- Potters, J. A., Curiel, I. J., & Tijs, S. H. (1992). Traveling salesman games. *Mathematical Programming*, 53, 199–211.
- Quade, D. (1979). Using weighted rankings in the analysis of complete blocks with additive block effects. *Journal of the American Statistical Association*, 74, 680–683.
- Qureshi, A. G., Taniguchi, E., & Yamada, T. (2010). Exact solution for the vehicle routing problem with semi soft time windows and its application. *Procedia-Social and Behavioral Sciences*, 2, 5931–5943.
- Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In 9th Python in science conference.
- Shapley, L. (1953). *A value for n-person games*. Cambridge: Cambridge University Press.

- Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*, 84–90.
- Stecca, G., & Kaihara, T. (2021). Negotiation based approach for collecting and recycling operations in circular economy. *Procedia CIRP*, *104*, 200–205. <https://doi.org/10.1016/j.procir.2021.11.034>
- Tan, S. Y., & Yeh, W. C. (2021). The vehicle routing problem: State-of-the-art classification and review. *Applied Sciences*, *11*, 10295.
- Terpilowski, M. A. (2019). Scikit-Posthocs: Pairwise multiple comparison tests in python. *Journal of Open Source Software*, *4*, 1169.
- Toth, P., & Vigo, D. (1998). Exact solution of the vehicle routing problem. *Fleet management and logistics* (pp. 1–31). Cham: Springer.
- Toth, P., & Vigo, D. (2002). *The vehicle routing problem*. Philadelphia: SIAM.
- Verdonck, L., Beullens, P., Caris, A., Ramaekers, K., & Janssens, G. K. (2016). Analysis of collaborative savings and cost allocation techniques for the cooperative carrier facility location problem. *Journal of the Operational Research Society*, *67*, 853–871.
- Young, H. P. (1985). *Cost allocation: Methods, principles, applications*. Amsterdam: North Holland Publishing Co.
- Yu, Y., Benjaafar, S., & Gerchak, Y. (2015). Capacity sharing and cost allocation among independent firms with congestion. *Production and Operations Management*, *24*, 1285–1310.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.